# What should the OSG look like in 2017?

- What value will OSG be providing to the scientific and computing communities?

- What products and services will OSG be providing?

- Which science communities will be using OSG, and in particular which of these are not currently using OSG?

- Who will be providing the computing resources?

- How will the OSG community be working together to advance their goals?

# Open Science Grid in 2017

**Thomas J. Hacker**

Associate Professor, Computer & Information Technology

Co-Leader for Information Technology, NSF George E. Brown Network for Earthquake Engineering Simulation

April 9, 2014

# What should the OSG look like in 2017?

# What value will OSG be providing to the scientific and computing communities?

- Continue to foster and build communities at the intersection of domain and computational sciences
- Which science domains would benefit from OSG?
  - Computational  (doing this well now)
  - Data Intensive (beyond ‘Big Data’)
    - Data collected from sensors in the virtual and physical worlds
- Building community, leveraging efforts to produce synthesis, sharing resources to maximize science return on investment
  - OSG has established a very successful model for how to build and sustain community driven infrastructure
  - Communities outside of Physics would benefit from OSG’s expertise and lessons learned

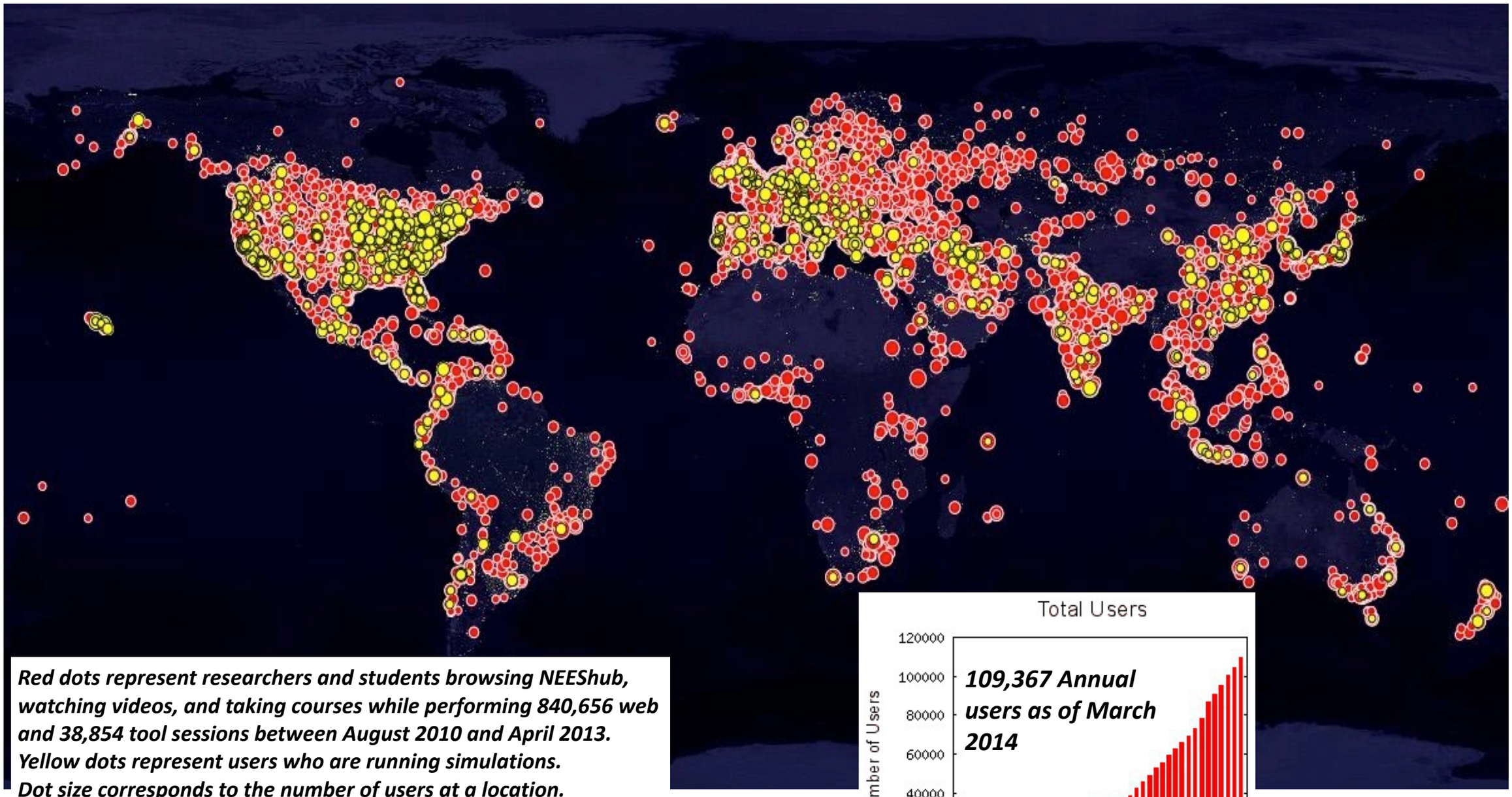
# What products and services will OSG be providing?

- Data needs of the science community that need to be addressed today and in the future
  - Data curation for long-term preservation of data to facilitate reuse
  - Protect and preserve university and community intellectual property
  - How will others be able to cite and reuse data 10 to 15 years from now?
  - Digital Object Identifiers (DOIs) for data
  - Establish authoritative scientific digital record
  - Support infrastructure for repeatable computational simulation and analysis
    - Verification and validation, and being able to handle uncertainty quantification
- Support the discovery of underlying models from observational and simulation data
  - Inverse models
  - Multi-dimensional uncertainty quantification and reduction of high dimension data to facilitate efficient exploration of high dimensional data
- Need for the preservation of software that can interpret these data
  - Virtual machines is a possible approach
  - Need to be able to run these software, not just archive them

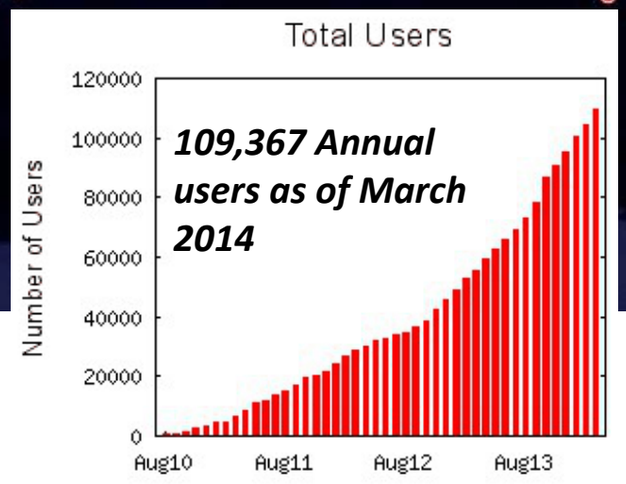
Which science communities will be using OSG, and in particular which of these are not currently using OSG?

- Computationally intensive science and engineering disciplines
  - Continue to support communities such as high energy Physics
  - High performance  / High throughput computing
  - Effectively use the most powerful computing capabilities

- Add capabilities to serve data intensive disciplines
  - Tremendous variety and increasing resolution of data from sensors
  - Example: Center for Urban Science + Progress at NYU
    - Keynote talk at SC13 about Urban Informatics
  - Communities increasing use of high performance / high throughput capabilities
  - Communities are large and international in scope



***Red dots represent researchers and students browsing NEEShub, watching videos, and taking courses while performing 840,656 web and 38,854 tool sessions between August 2010 and April 2013.***
***Yellow dots represent users who are running simulations.***
***Dot size corresponds to the number of users at a location.***



# Who will be providing the computing resources?

- Branscomb pyramid
  - Amazon EC2 and cloud providers will likely become a viable alternative at the lower layers of the pyramid
  - Which layer(s) will OSG seek to address?
  - Universities will continue to seek to meet the needs of students, faculty, etc.
  - Universities will likely seek to establish consortiums to minimize costs and spread the risks
- What about Data?
  - Few good institutional models for providing infrastructure for data
  - NEES (today)
    - Over 2M project files and directories from the earthquake engineering community
  - University of Michigan Institutional File System (IFS) project (1990's)
    - "One world, one filesystem"

# How will the OSG community be working together to advance their goals?

- OSG has developed effective governance, sharing, and scaling models
- Who?
  - Which domain science and computational communities will the OSG seek to serve?
- Where?
  - National or international?
- What?
  - What kinds of services will the OSG seek to provide?
- When?
  - Over what timeframe will these needs emerge, and when will OSG be able to address them?
- **This all leads to how**

# The OSG in 2017?

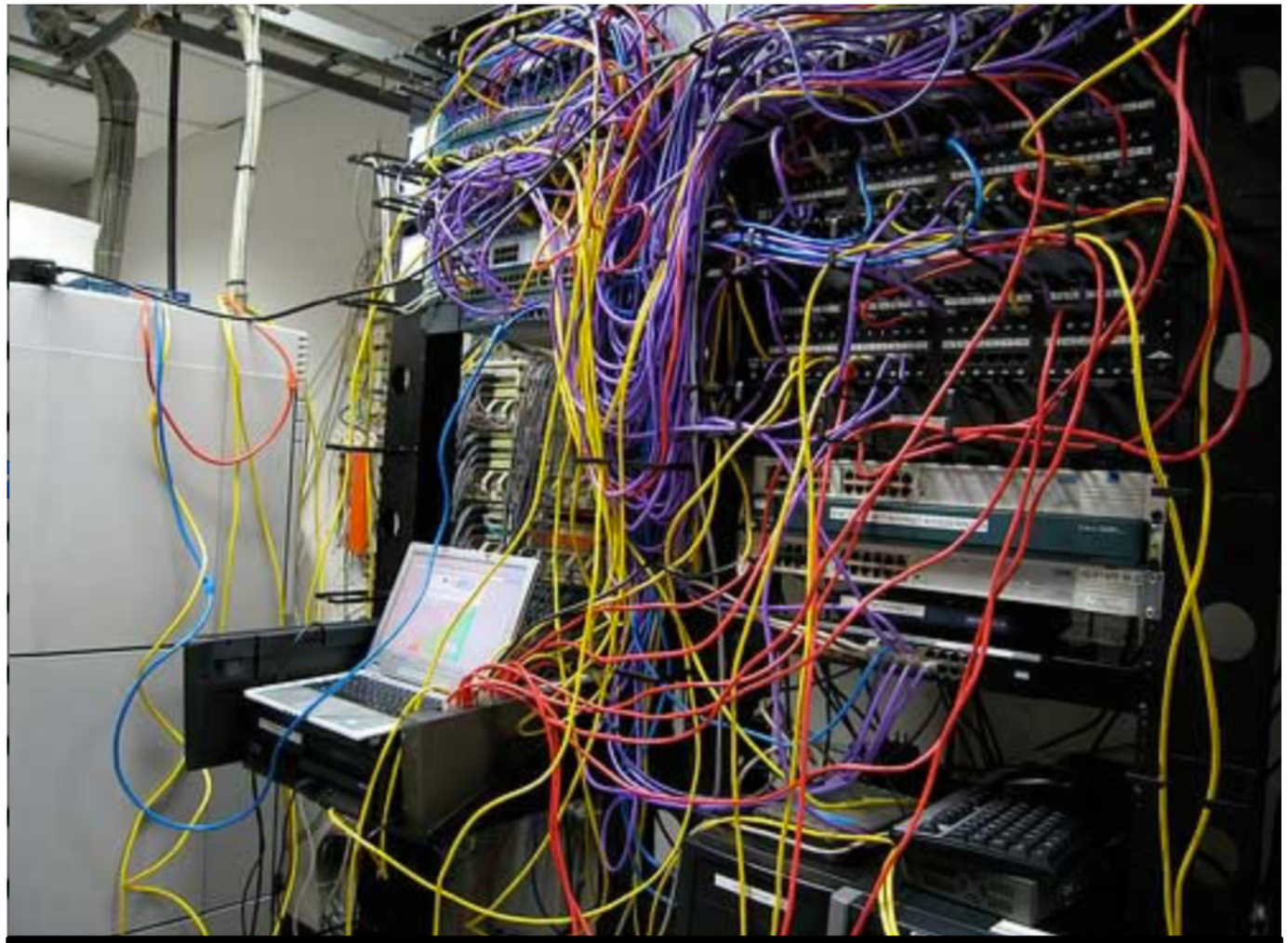
Discovery
Unconstrained by
Geography

**April 9, 2014**

**Brooklin Gore**

# Challenges to Network Adoption

- Causes of performance issues are complicated for users.
- Lack of communication and collaboration between the CIO's office and researchers on campus.
- Lack of IT expertise within a science collaboration or experimental facility
- User's performance expectations are low ("The network is too slow", "I tried it and it didn't work").
- Cultural change is hard ("we've always shipped disks!").
- Scientists want to do science not IT support



**The Capability Gap**

Scientist
+
Data
+
Network
+
Computing

Higgs
&
Cancer
&
Energy
&
Climate

# ESnet Science Engagement Services

**Partnerships**

With facilities / research teams / providers, building foundation for lasting impact.

**Education & Consulting**

Webinars, workshops, 1:1 data mobility consultations with scientists, support teams.

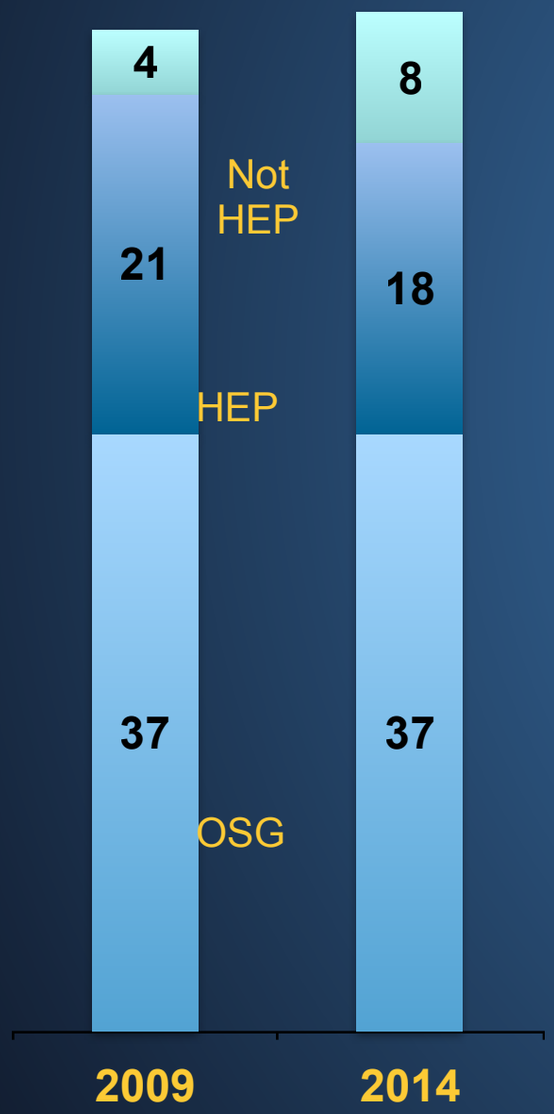
**Resources & Knowledgebase**

Reference designs, case studies, papers, FAQs – tailored for multiple audiences.

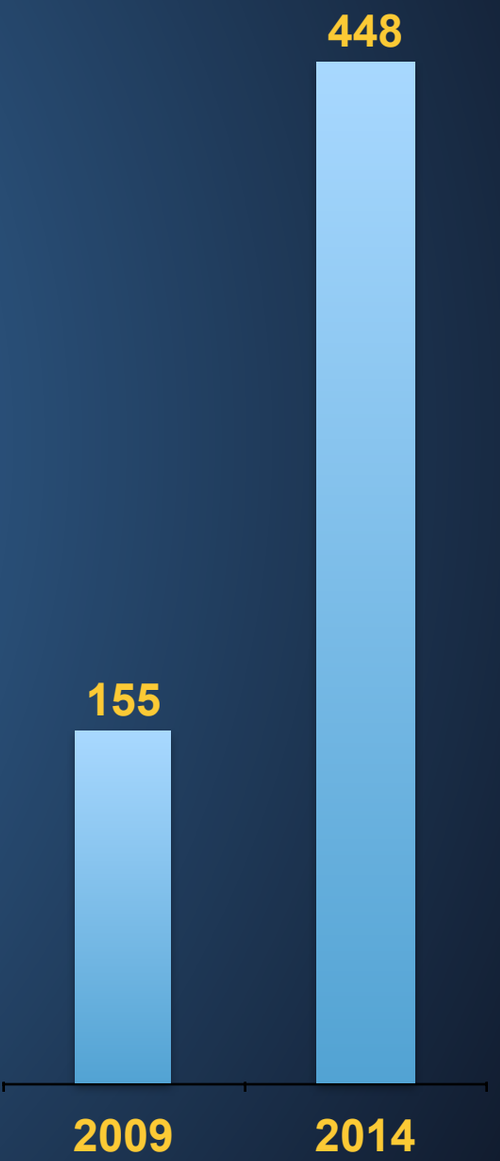
We built this really complex computing thing. It works! Some scientists are actually using it. - 2009

We built this really complex computing thing. It works! Some scientists are actually using it. - 2009

**OSG Annual Report Pages**

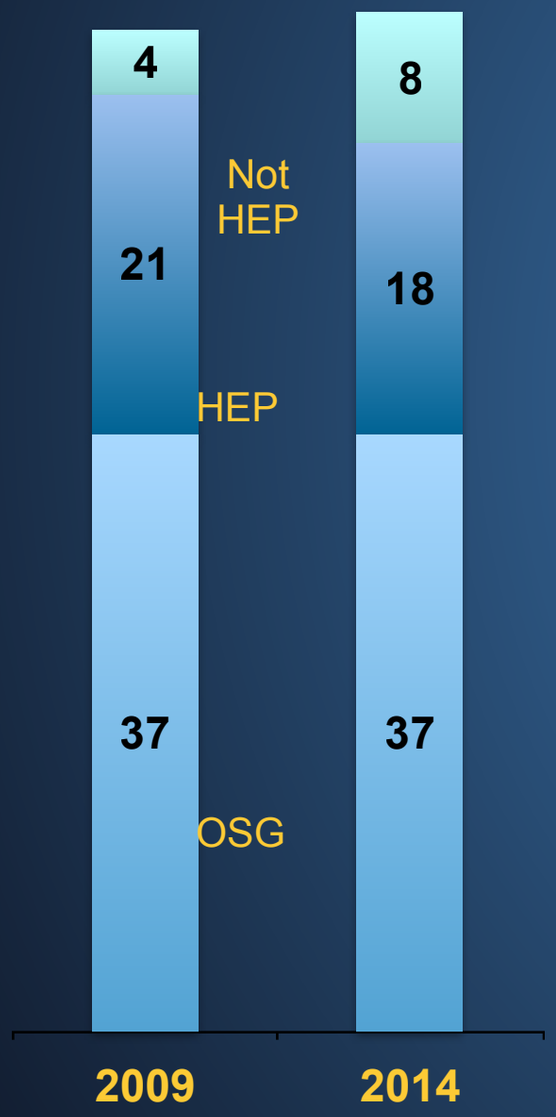



**OSG Citations**

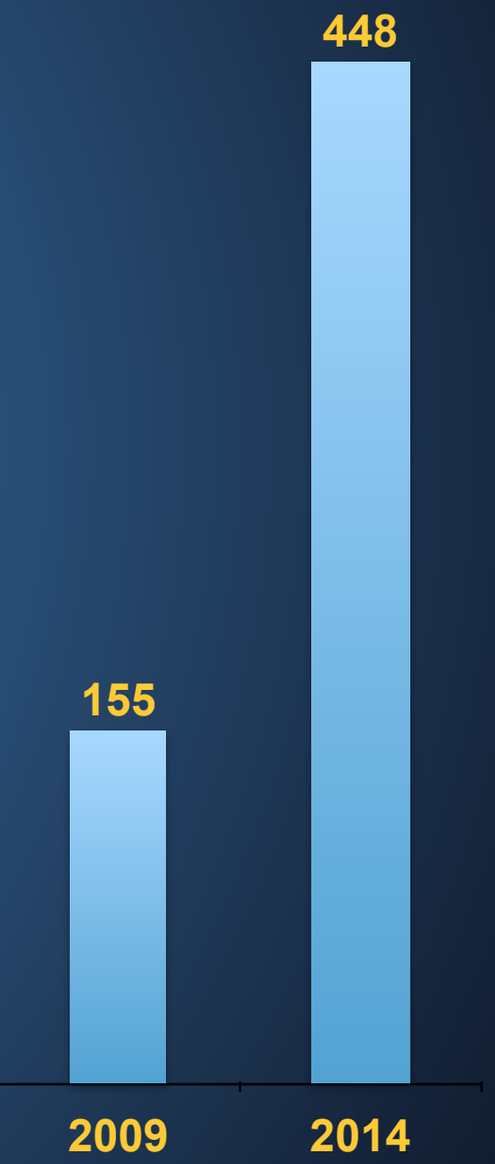


We built this really complex computing thing. It works! Some scientists are actually using it. - 2009

**OSG Annual Report Pages**




**OSG Citations**




Citations drop dramatically…everyone just does it this way now… - 2018

# Fkw's perspective on OSG in 2017

- On- and off-ramping to/from National CI will be common among major Research Institutions (Universities, National Labs, Foundations)
  - Seemless, policy driven, under the control of the "scientist as enduser"
- Divergence of HTC into heavy CPU (co-processor dominated) and heavy IO (BigData "filtering")
  - Metascheduling of clusters w. co-processor capabilities.
- Everything is less static
  - Allocations & provisioning & placement & access
  - Mixed commercial, public & private infrastructure

# OSG in 3 Years

Ian Fisk
April 9, 2014

# Evolving from

- It's the "Open" Science Grid, it sometimes feels we concentrate too much on whether our largest stakeholders are happy
- Maybe it would be OK for OSG to have a unique identity and mission of its own
  - We spend a lot of effort dragging legacy services with us that we designed for LHC ten years ago
    - CE, SE, central Information services, certificates are all in some way deeply flawed. How do we get out?

# Evolving to?

- We should be working toward a much less static system. Resources providers and resource consumers should be continuously flowing in and out
  - We have tools for taking the environment (cvmfs + parrot, VMs)
  - We have tools for lighter weight resource provisioning (BOSCO)
  - We have tools to deliver data from a distance (Data Federations over Xrootd)

# The Future

- Open Science Grid will continue to be useful but won't grow only by being better for LHC.
  - LHC needs opportunistic computing and has an expected resource growth for a program that lasts 20 more years
- OSG can evolve in influence by being indispensable to many smaller sciences
- Goals: Need to measure resource registration and application adoption in hours
  - Last year we provided 30M CPU hours over the year. Imagine 100 small sciences, with 1000 core problems, it's 12 days of processing
    - Adoption time should be less than 10% of usage time, so you need to be up in about a day

# Emphasis

- Our emphasis should be on
  - Fast and transient resource contribution and provisioning
    - Authentication management
    - Providing data and dynamic storage
- OSG Connect is the right direction, but should be established as the strategic direction of OSG